The studying visual-based navigational guidance with network-based agent

Abstract

Social insects such as wasps and ants are central-place foragers, which means they perform frequent round-trips between the central-place and the place for foraging. In order to return to the nest from an arbitrary position in the surrounding scene, the insects capture the spatial landmark features in the surrounding scene with their visual sensory receptor and correlate these features with the path to the nest. Although such mechanism of landmark capture has been observed through behavioral experiments of insect, there is few work determines what and how the spatial landmarks in real environment contribute to visual navigation. Since some insects can navigate to their nest using only panoramic vision, we applied convolutional neural network for extracting the landmarks in visual fields. Our work contributes to a network-based agent that utilizes the panoramic vision to guide its motion for imitating the motion of real wasps. With our visualization of external motion behavior and stimulated internal neural activation, we find the visual landmarks, in the form of activated neurons in the visual field, are able to directly guide the stimulated wasp in a 3D model scene, while the tree line and skyline are more likely to be the landmark for wasps.

Introduction

It has been found that some specific social insects have the ability to learn the foraging routes with pure visual information, and the ability is supported by their visual sensory receptors and varying navigational strategies [1, 2]. Many comparative experiments have been conducted for proven that some insects, such as ants and wasps, apply memorized landmark to guide their path home [3, 4]. Meanwhile, due to the lack knowledge of insect brain, how the memorized landmarks help for homing navigation through neural processing is still a mystery for us.

In this case, we explore a model of brain neurons to construct an agent for mimicking the insect motion against varying visual input. Benefit from the development of convolutional neural network algorithm, a network-guided agent is able to mimic real insect behavior with this biologically inspired algorithm. After feeding the records of the behaviors of different insect species, it is feasible to conduct repeated experiment on the stimulated agent for obtaining general observation with little individual bias. At the same time, the visualization of the neuron activation inside of the artificial network can unveil the information processing, which provides a new view on how specific spatial features in the visual field contributes to biological motion decision.

In this paper, we develop a convolutional neural network based visual navigator by training it with wasp learning flights and the views within a 3D reconstruction of their natural environment. Our study of vision and motion is based on the 3D model constructed and recorded wasp foraging paths contributed by Murray and Zeil [6, 7]. At the same time, we apply weight regularization to highlight activated neurons for capturing the critical pixel-wise landmarks in the visual field.

For the experimental details, we visualize densely distributed motion in form of vector of the stimulated network-guided agent with different starting states. Such visualization consists of two ranges of starting locations, one is near the nest and the other is distant to the nest, while different orientations are involved as a variable for controlling the starting state of the agent. At the same time, the neural activation at different locations and orientation are compared for determining the influence of spatial feature on motion decision.

As for the result, we find our stimulated agent is able to approve the nest with some certain starting states. There is also relationship existing between neuron activation and corresponding predicted motion, while such relationship is still ambiguous based on our analysis. At the same time, we notice the skyline and tree line usually play a role as the landmarks during the homing path of the agent.

Material and methods

Model and path data for network training

We train the convolutional neural network model using Murray & Zeil's reconstructed Mount Majura 3D model [6; Figure 1]. The central region in the environment is a flatland that is surrounded by trees with different appearance features. The nest is set between two poles as the landmark in the center of the scene. The model is rendered using Unit3D (http://unity3d.com, v5.4), which renders panoramic views from user-defined locations and orientations using three virtual cameras. Images are output in PNG file format.

Figure 1: Left: 3D model of the wasps nest clear on Unity (Mt Majura Canberra), Right: the scene around the nest in the 3D model.

Figure 2: The $360°$ panoramic view from the position of the nest

[6] provides the measurements of wasp foraging flight path within the Mount Majura in Canberra, these measurements record the absolute position as well as the absolute orientation of flying wasps in continuous timestamps. We select five flight paths as the training set for our model.

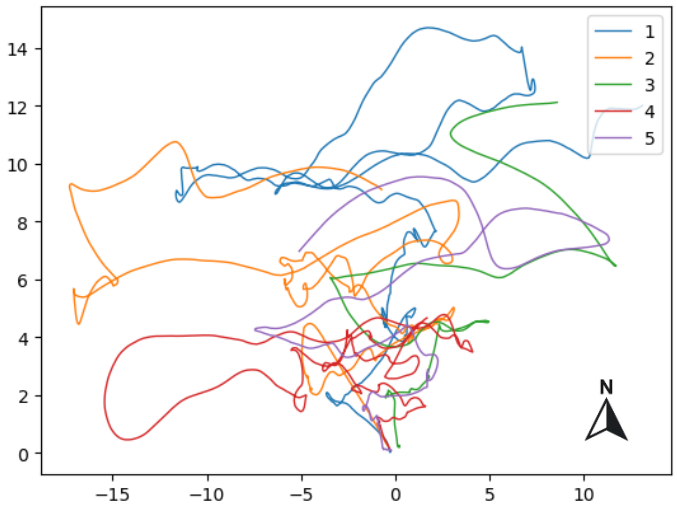


Figure 3: The Illustration of the flight paths used for training. As foraging flight paths, all five trajectories start around the nest $(0,0)$ and terminate at different foraging spots in the scene.

For expressing the state of a wasp in the t^{th} timestamp, we denote the state with $S_t \in \mathcal{R}^4$ that includes the absolute positions and absolute orientation. Such data has a format as below.

Table 1: A example of path data, $S_t|_{x,y,z}$has unit meter, while $S_t|_{angle}$ has unit degree.

Index	$S_t\|_x$	$S_t\|_y$	$S_t\|_z$	$S_t\|_{angle}$
1	-0.122760	0.448258	0.086437	110.122671
2	-0.141913	0.469608	0.057027	110.122671
…	…	…	…	…

It is noted that S_t directly controls the vision field. In another words, each vision field of the agent in the 3D model with S_t is unique. We captured the left, front and right RGB camera view $\{im1_{S_t}, im2_{S_t}, im3_{S_t}\}$ for each state S_t for simulating the panoramic vision of a real wasp.

In order to train our model to approach a consistent termination point as nest, we took the reversed foraging paths for imitating homing paths. In this case, the training paths start from arbitrary locations in the scene and terminate in the nest. Since these stimulated homing paths have a common target that is the nest, we expect the network-guided agent learns to correlate the visual information and the motion for approving the nest after training.

Network-guided agent

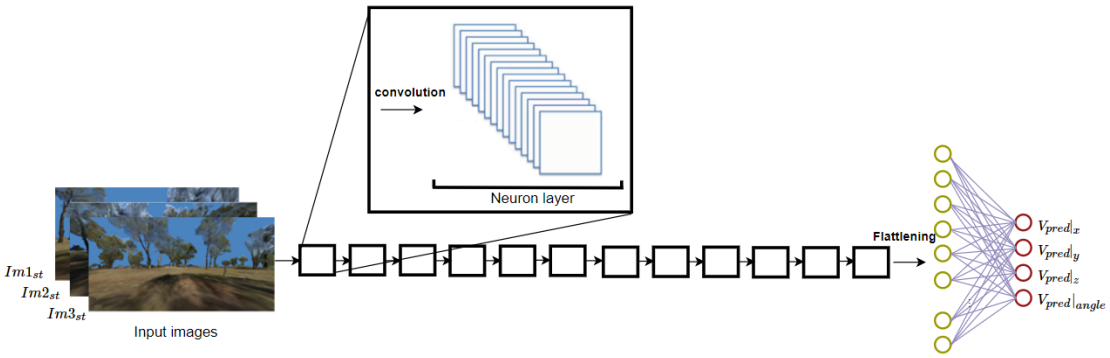


Figure 4: The Illustration of network that guides the motion of the agent. It takes the panoramic images $\{im1_{S_t}, im2_{S_t}, im3_{S_t}\}$ as input. By applying convolution operation, the network with 11 cascaded neuron layers is able to predict the instantaneous relative motion $V_{t_{pred}}$

Given several sequences of homing paths and the corresponding images in visual field, we designed an 11-layer convolutional network for guiding a wasp agent to learning the general relationship between the homing motion and the vision with the training flight paths.

To enable a network learning the long-range homing flight path, we express the motion from the start point S_1 to the nest S_T with consecutive instantaneous relative motion V_t. Here, V_t is in the form of vector and defined by the difference between two continuous state S_{t+1} and S_t.

$$V_t = \{S_{t+1}\} - \{S_t\}$$

where $t \in [1, \dots, \mathrm{T}]$.

The stimulated agent is guided by an 11-layer conventional neural network (CNN) that take panoramic images $\{im1_{S_t}, im2_{S_t}, im3_{S_t}\}$ for predicting the instantaneous relative motion $\{V_{t_{pred}}\}$.

During training period, we adjusted the neural weight in the network for minimizing the simulating error between the predicting motion and real predicting motion across the homing paths. The error is defined by

$$\begin{aligned} e_{stimu} = &\sum |S_{t+1}|_x - (V_{t_{pred}}|_x + S_t|_x)| \\ &+ \sum |S_{t+1}|_y - (V_{t_{pred}}|_y + S_t|_y)| \\ &+ \sum |S_{t+1}|_z - (V_{t_{pred}}|_z + S_t|_z)| \\ &+ \sum \left| S_{t+1}|_{angle} - \left(V_{t_{pred}}|_{angle} + S_t|_{angle}\right)\right| \end{aligned}$$

By minimizing the simulation loss, the agent is expected to learn the similar navigation behaviour as wasps conducted in real life.

Neural activation of attention

Similar mechanism of landmark representation has been found in both the artificial CNN and the biological vision system [8]. The previous works used to roughly determine the landmarks by

statistically analyzing the objects in the visual field of a biological creature, while the neural activation of the artificial CNN provides an approach for visualizing the landmarks through the information flow transmitted between neurons. In this paper, we recognize the critical spatial landmark that contributes to instantaneous motion $V_{t_{pred}}$ by visualizing the activated processing neurons.

The technique of weight regularization has been proven that is able to result in a visually observable neural activation [9]. It works as penalty term that is summed up with the stimulation error e_{stimu}. For acquiring a sparse-distributed pixels, we apply L1 regularization that is defined as

$$L_1 = \sum a||w||_1$$

where w represents the neural weight, while a is a variable that controls the proportion of penalty. In our experiment, we implicitly set $a = 0.001$, which is able to illustrate a clear neural activation map.

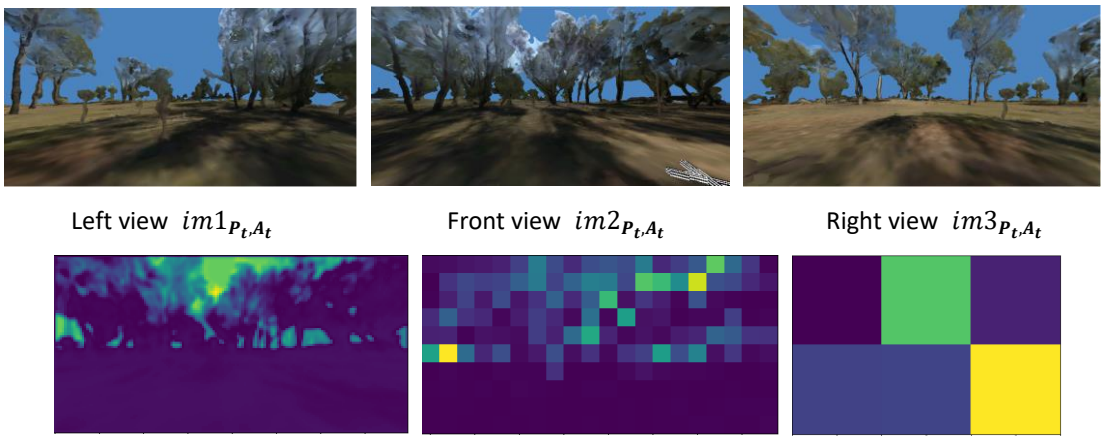


Figure 5 The demonstration of neural activation of a panoramic vision. **Top:** vision input $\{im1_{P_t,A_t}, im2_{P_t,A_t}, im3_{P_t,A_t}\}$, **Bottom:** The corresponding neural activation map in the 3^{rd}, 6^{th} and 10^{th} layer

With a panoramic image as input, our trained CNN model generates the instantaneous motion $V_{t_{pred}}$ as well as the corresponding feature maps in a compressed size.

The highlighted values in the feature maps represent the activated neurons, and it indicates these neurons transmit a larger amount of information compared with others. Since the neurons in the neighboring layers are locally connected, the activation is locally transmitted, which means the neuron activation in the first several features maps is able to reflect the attentions in the input image.

For the instance in Fig. 5, the neural activation map compresses the spatial information in the front view $im1_{P_t,A_t}$, the left view $im2_{P_t,A_t}$and the right view $im3_{P_t,A_t}$. We can observe that the pixel-wise neuron activation of the 3^{rd} layer occupies the tree crown in the front view $im2_{P_t,A_t}$, and it indicates this region has the highest contribution for guiding our agent to conduct the next motion. The example of processing in the CNN is shown in Appendix.

Experimental setting

Considering the 3D model provides with complex spatial features, it is hard to explicitly figure out the how vision helps our network-guided agent approach a specific site. In this case, we heuristically set multiple spatially dense starting states $S|_{x,y,z,angle}$ for the agent and observe its performance.

The spatial location $S|_{x,y,z}$ of the starting state are selected in two spatial density manners with respect to location of two poles around the nest. In this case, we assumed the height of all starting point is 10 centimeters above the ground.

For the near-nest, we selected 113 different starting points between the two poles. As the left image shown in Fig. 6, the starting points are evenly distributed between the poles (marked as red dots). As for the long-range, the positions of 256 points are illustrated in the right image in Fig. 6, which are evenly distributed in a long-range around the two poles.

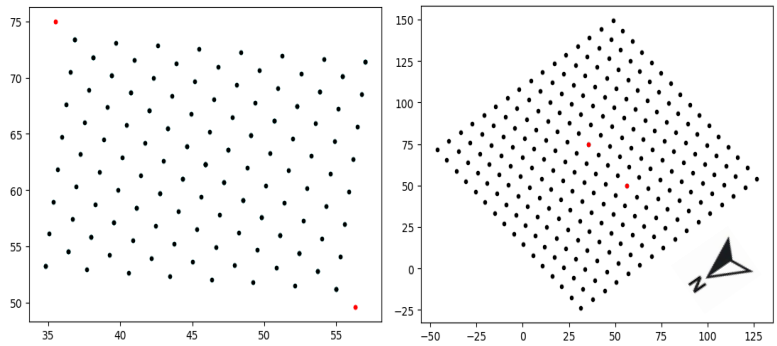


Figure 6: Positions of the starting point for short-range and long-range experiments

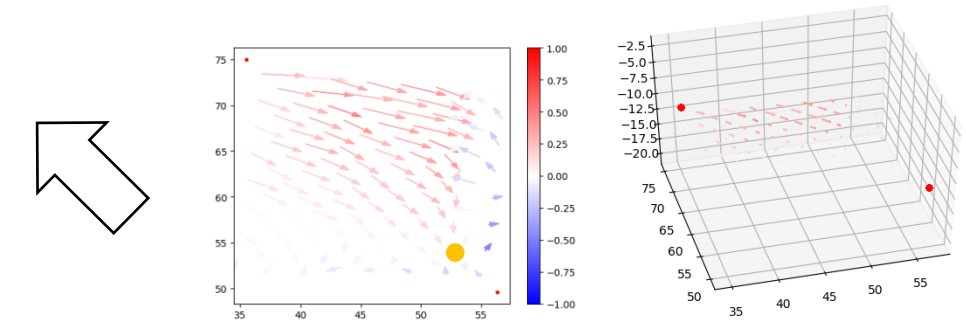
In each experiment, all starting states should have uniform orientations for controlling variables. The orientation of the state $S|_{angle}$ is set as $40°$, $130°$, $220°$ and $310°$, which represents the east, the north, the west and the south respectively. With two sets of $S|_{x,y,z}$ and four sets of $S|_{angle}$, our experiment includes eight settings of the starting states $S|_{x,y,z,angle}$.

Experimental Results

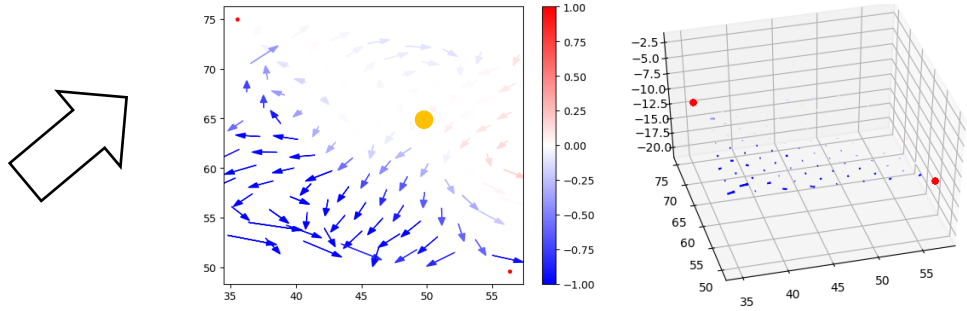
For acquiring an intuitive visualization of the stimulated motions, we employ arrow-style vectors for representing the motion magnitude and the motion direction parallel to the horizon, while the color of each vector indicates the corresponding steering motion $V_{pred}|_{angle}$. We construct the two-dimensional visualization and three-dimensional visualization for $V_{pred}|_{x,y}$ and $V_{pred}|_{x,y,z}$ respectively.

Near-nest agent navigation paths

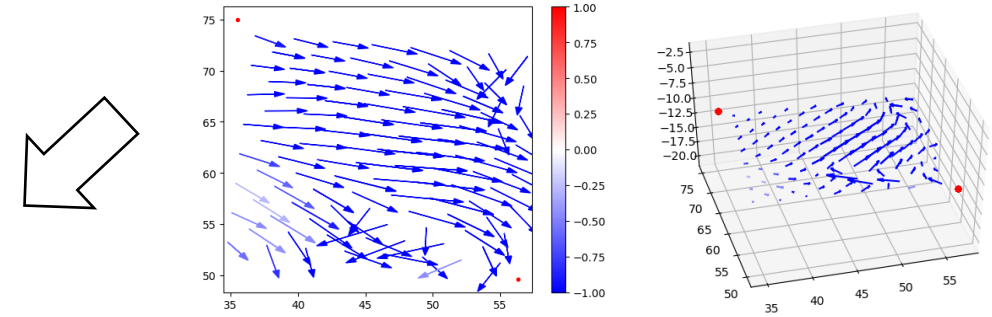
a Motion visualization when $S|_{angle} = 40°$



b Motion visualization when $S|_{angle} = 310°$



c Motion visualization when $S|_{angle} = 130°$



d Motion visualization when $S|_{angle} = 220°$

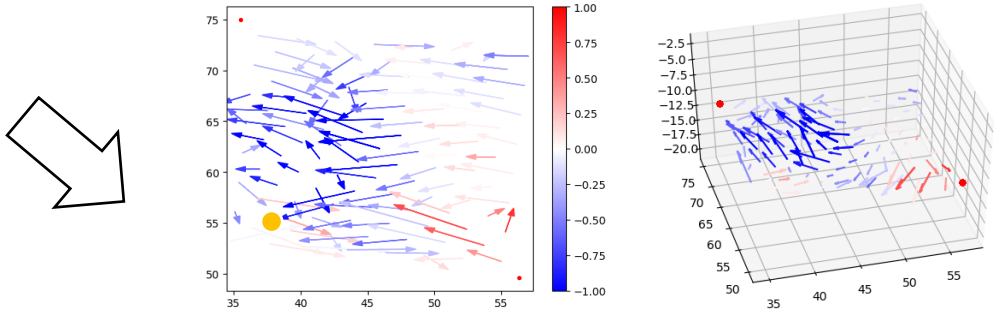


Figure 7: visualization of predicted instantaneous motion V_{pred} with the four different uniform starting points in two dimensions and three dimensions. Each arrow-style vector represents a predicted motion $V_{pred}|_{x,y,z}$ in the corresponding directions, while the magnitude of motion is reflected by the length of the vector. The color of vector shows the magnitude and direction of steering motion $V_{pred}|_{angle}$. (Red is leftward and blue is rightward steering motion).

Left: The arrow indicates the orientation of uniform starting points. **Middle:** The motion $V_{pred}|_{x,y}$ in two dimensions space. **Right:** The motion $V_{pred}|_{x,y,z}$ in three dimensions space.

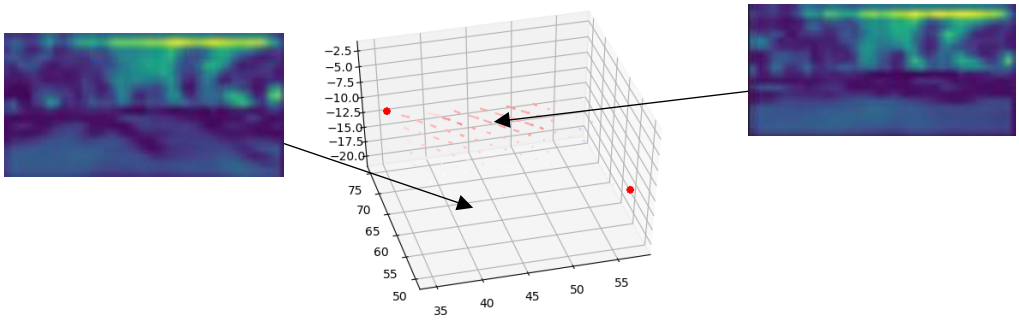
With Fig. 7a and Fig. 7b, we can find the arrow-style vectors in these two graphs have a uniform direction and toward fixed regions with a decreasing magnitude of motion. Such regions (marked with yellow dots) are surrounded by motions with relatively small magnitude, which indicate such regions can be considered as the exact termination for the homing path for $S|_{angle} = 40°$ and $S|_{angle} = 130°$.

As for Fig. 7c, it shows more significant motions compare to the Fig. 7a and Fig. 7b. The motion vectors in Fig. 7c have large magnitude of motions and an obvious steering motion and for turning rightward. By considering the direction of starting state, $S|_{angle} = 130°$ is opposite to the direction of training paths that is moving toward the south (shown in Fig. 3), although the magnitude of motion does not decline as they approach the nest locates in $(50,65)$, the direction of the motion vectors change. It means the agent makes a correct inference of its motion without directly guidance of training paths. At the time, since the agent only learns to approach the nest from the north, it reserves a rapid motion even when it is close to the nest.

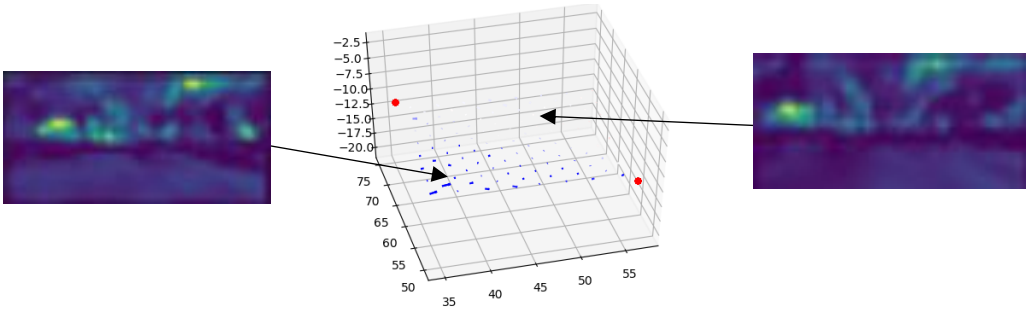
Fig. 7d presents a noisy pattern of motion vectors compared with the other treatments. Since the near-nest region does not provides a significantly different visual field, it means the subtle change of visual information leads to obviously different motions.

Visual-based analysis of the near-nest experiment

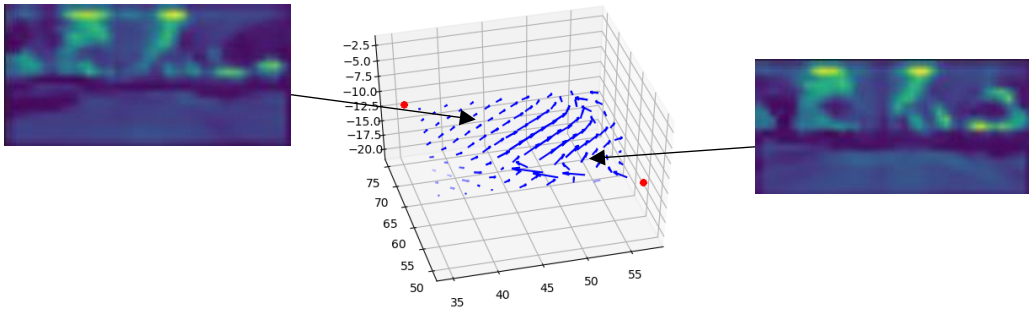
a Motion visualization when $S|_{angle} = 40°$



b Motion visualization when $S|_{angle} = 310°$



c Motion visualization when $S|_{angle} = 130°$



d Motion visualization when $S|_{angle} = 220°$

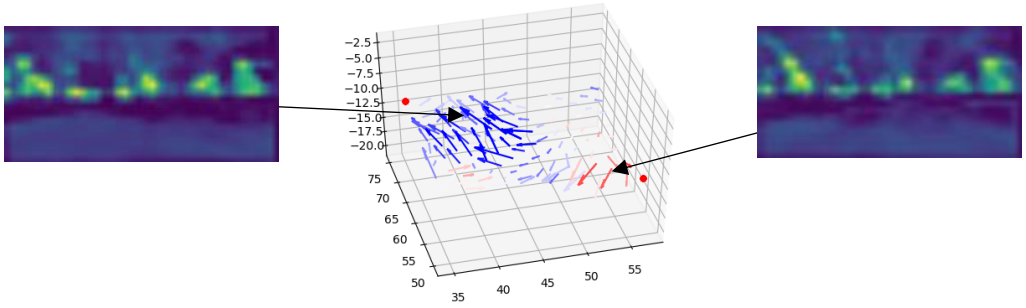


Figure 8: Three-dimensional motion as well as its neuron activation for near-nest navigation

The comparisons of neuron activation in Fig. 8b and Fig. 8d internally expose the activated spatial features for different motion behaviors.

When $S|_{angle} = 220°$, the two clusters show the similar patterns of neuron activation, while both neuron activation are highlighted in the middle region of map. By comparing the activation maps in Fig. 8, we can observe that there are more separated activated neuron clusters in Fig. 8d compared with others. By considering such separated activation as the landmarks, we could infer the agent starts with $S|_{angle} = 220°$ is guided by multiple landmarks, which make the agent hard to conduct a consistent motion with different starting point.

For the case with $S|_{angle} = 310°$ in Fig. 8b, the right neuron activation map has less activated neurons then the left activation map, the right corresponding motion also has a smaller magnitude compared with the motion in the left. Meanwhile, in the case with $S|_{angle} = 310°$, Fig. 8d also presents two motions with different magnitudes, but their neural activations look similar. It leads to ambiguous relationship between the activation of neurons and the predicted motion, which is hard to be clarified in this experiment.

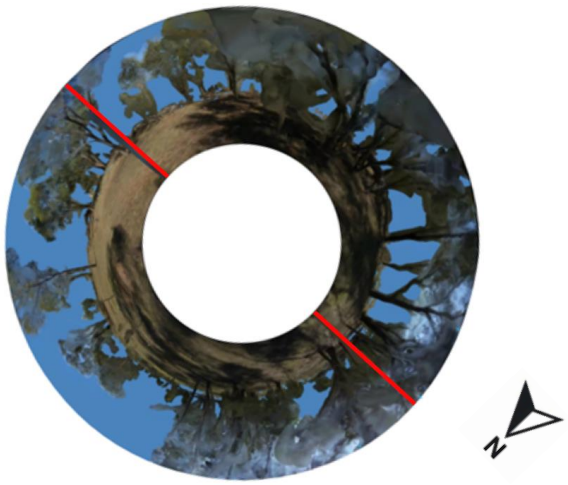


Figure 9: Reference $360°$ panoramic vision around the nest (Red line indicates the poles in the corresponding direction)

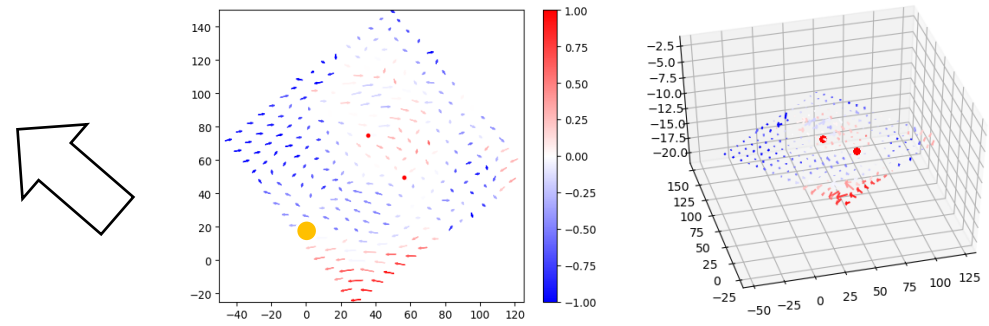
By annotating the position of poles on Fig. 2, the panoramic view shown in Fig. 9 is able to help bridge the predicted motion with the spatial features.

By comparing Fig. 8 and Fig. 9, we find the skyline in the east, the tree line in the north and the west are the most important visual scene features for guiding navigation by visualizing the neuron activation map and identifying critical features (for process, see Appendix 1). Meanwhile, it is noted that the orientation of west shows noisy motion map in Fig. 8d, which is related to the dense trees in the west of Fig. 9. In this case, we suppose these dense trees provides dense spatial features that misleads the agent.

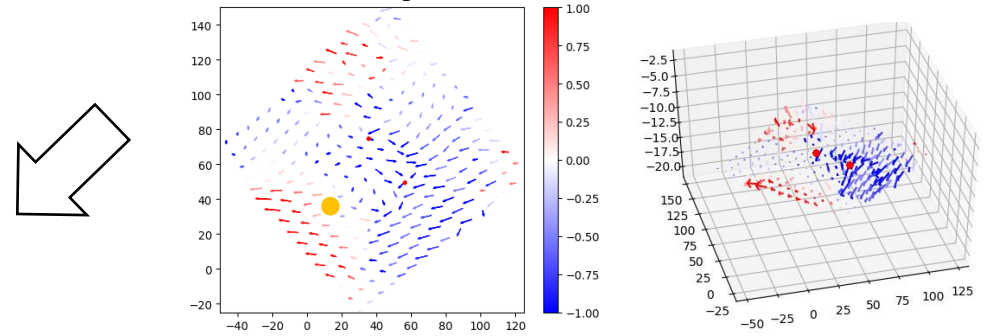
Long-range agent navigation paths

Differ from the motion visualization in the short-range experiment, the long-range experiment provides a significantly varying spatial information in the visual field for the agent, and it contributes to a generalization of the motion of stimulated agent as the starting points spread across the flatland in the 3D model, which is outside the bounds of training flight paths.

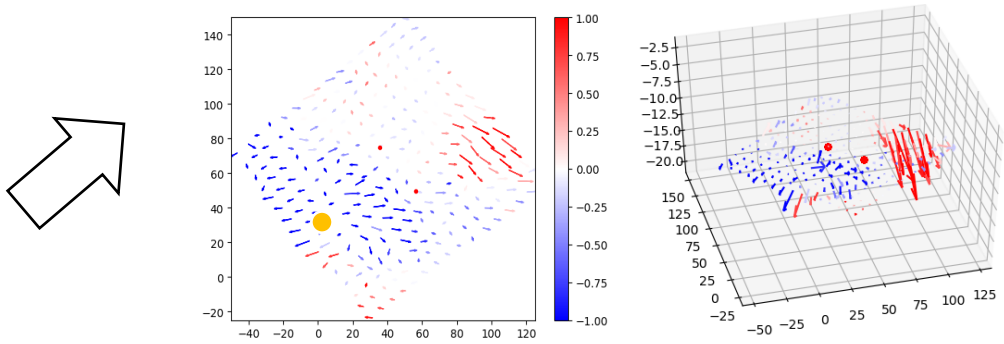
a Motion visualization when $S|_{angle} = 40°$



b Motion visualization when $S|_{angle} = 130°$



c Motion visualization when $S|_{angle} = 310°$



d Motion visualization when $S|_{angle} = 220°$

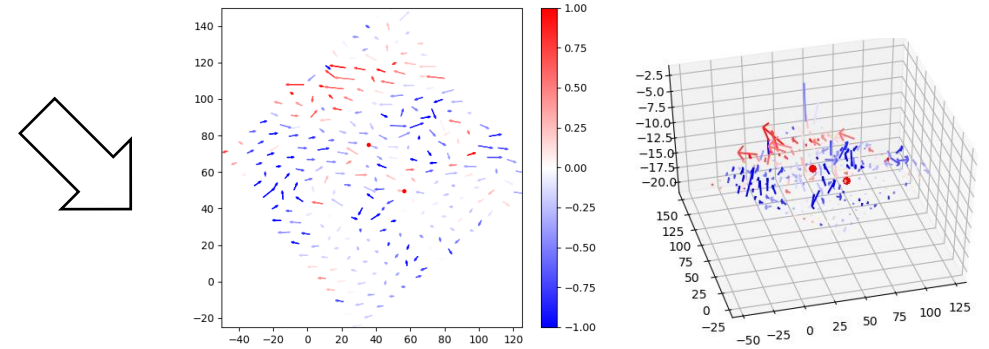


Figure 10: visualization of predicted instantaneous motion V_{pred} with the four different uniform starting points in two dimensions and three dimensions. **Left:** The arrow indicates the orientation of uniform starting points. **Middle:** The motion $V_{pred}|_{x,y}$ in two dimensions space. **Right:** The motion $V_{pred}|_{x,y,z}$ in three dimensions space.

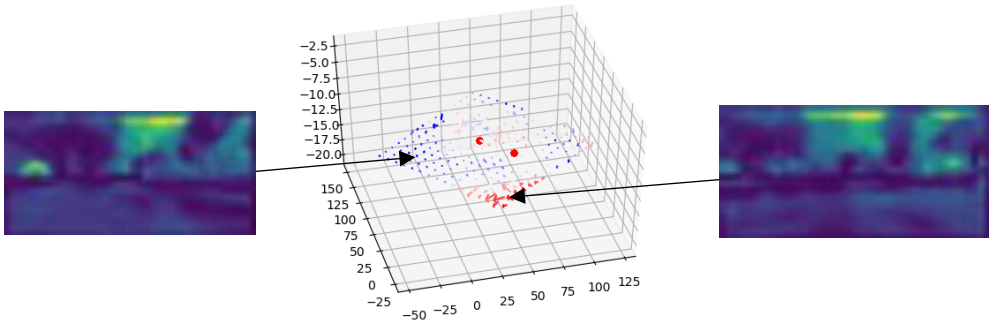
In Fig. 10, the four motion visualization maps consists of vectors with varying magnitudes and varying directions, while it is noted that the change of motion across a map is still in a gradual manner with respect to their starting position, and it means the control of the visual field is guiding the predicted motion. Meanwhile, there is few of commonality of the either the magnitude or direction exists among all these maps.

Although we manually set the point between the two poles as the termination in the training flight paths, in Fig. 10a, Fig. 10b and Fig. 10c, we still observe a termination region at the lower left of the two poles has slight motion around it (marked with yellow dots). It means the agent does not precisely learn the nest as its termination.

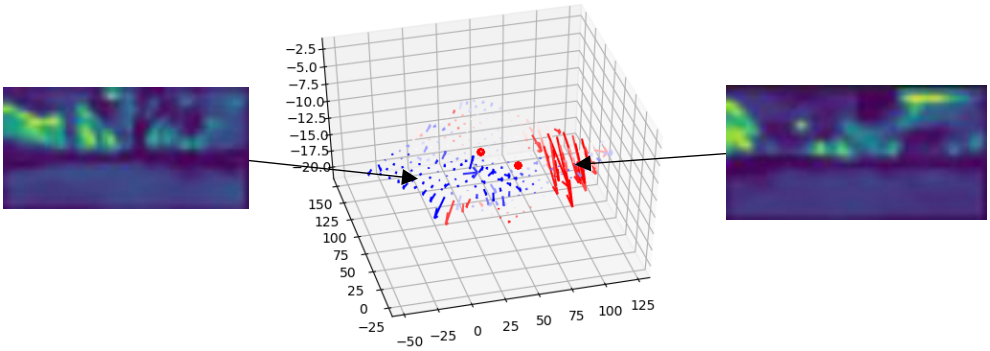
At the same time, the motion with state angle $S|_{angle} = 220°$ shows an irregular pattern compared with others, which is locally inconsistent and is similar to near-nest condition. It supports our conjecture in the previous section that some spatial features in the visual field are ambiguous for guiding the agent.

Visual-based analysis of the long-range experiment

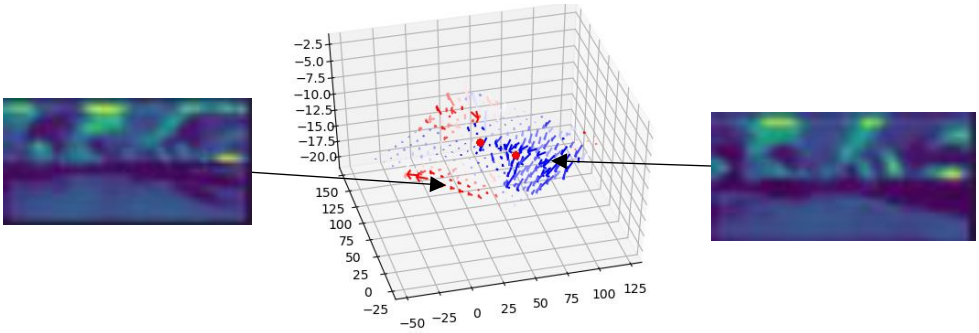
a Motion visualization when $S|_{angle} = 40°$



b Motion visualization when $S|_{angle} = 310°$



c Motion visualization when $S|_{angle} = 130°$



d Motion visualization when $S|_{angle} = 220°$

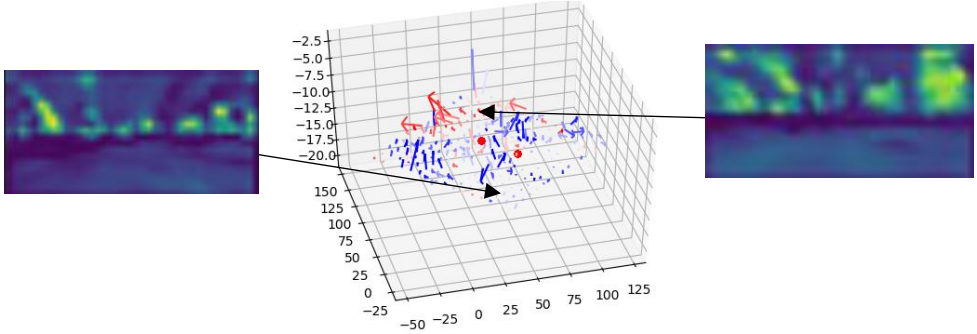


Figure 11: Three-dimensional motion as well as its neuron activation for long-range navigation

The neuron activation maps in Fig. 11 is based on a significantly changed visual field, it also results in varying neuron activations.

We can find the two pairs of motion vectors that pointed by arrows have different magnitude and direction in Fig. 11 and Fig. 11; however, the corresponding neuron activation pairs in Fig. 11 and Fig. 11 are in contrary manners, one pair looks different while another is similar.

This inconsistency between the neuron activation and the motion behaviour leads to a conclusion that the visual-based guidance is a complex processing that cannot be explicitly exposed in our experiment.

Statistical summary of neuron activation

Through all neuron activation maps, we notice most of the highlighted regions are in the upper half part. For validating this observation, we calculate the sum all the neuron activation of the visual images appear in the training paths, and plot the average magnitude of neuron weight verse their distance to the top of visual field.

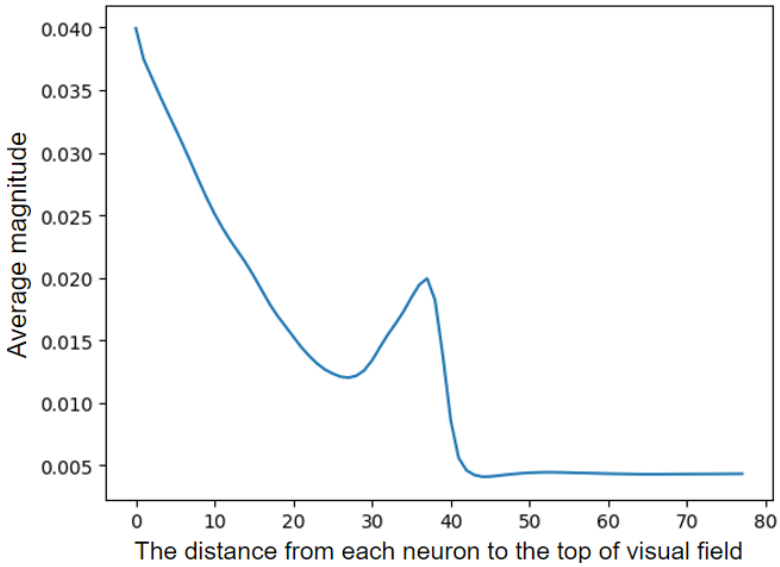


Figure 12: The average neuron activation across all training samples

With the plot in Fig. 12 we can find the neurons locate in the bottom half region always has less magnitude compared with upper half region, which indicates the ground is almost ignorable during homing navigation for our trained agent. Also, there are two peaks of the magnitude occur with the distances 0 and 38, such peaks represent the region of skyline and tree line respectively, which implies such region has a higher contribution for the motion.

Conclusion

As a conclusion, we apply convolution neural network to design a network-based agent. This agent is trained using 3D Unity model with the supervision of recorded wasp flight paths and

corresponding views for learning the behaviour of real wasps. With the behavioural experimental result, we find our agent is able to utilize the visual information as guidance to approve the nest with some certain starting states. Meanwhile the agent cannot precisely terminate at the nest if it starts from a spot that distant to the nest.

As for the studies of the neurons inside the network, we apply L1 weight regularization for obtaining visible neuron activation, and the generated neuron activation map unveils the landmark that guide the motion of agent. We find there is relationship existing between the landmarks and motions, but such relationship is still ambiguous with our experiments. Meanwhile, with statistical analysis, we conclude that most of the neuron activations appear in the region of tree line and skyline.

The case with an initial orientation of $220°$ is special as it shows an inconsistent motion in local region, which is abnormal compared with other conditions. For its neuron activation map, it shows the agent is guided with multiple landmarks. In this case, its motion is sensitive to the change of visual field, as all the landmarks will change with the visual field at the same time. By observing the objects in the visual field, we suppose the dense trees cause such problem.

Overall, the applying of the network-guided agent supports repeated experiments for generating a large number of motion vectors that forms a motion field, while such experiment is hard to be implemented on living creatures with no individual bias. The biologically inspired model also provides a baseline stimulation of visual-based navigational system, while the visualization of neurons and the functionality of neuron layer could help to explore the biological neural processing.

Our current experimental result is lack of quantitative analysis, and it is limited by the acquisition of annotations. For instances, if the annotation of the objects in the visual field is available, we can acquire the 'object-wise activation' based on neuron activation and it can lead to a statistical analysis on determining which kind of object provides the most significant guidance to the wasp.

As for future works, different architecture of network can be applied for guiding the insect agent. For instances, spiking neuron network (SNN) has the mechanism of membrane potential that is more closely mimic natural neural network. We can expect such network architecture is able to demonstrate a different neuron activation behaviour. Also, it is worth generating larger datasets of insect behaviour, which can contribute to a trained agent with less Individual bias.

Reference:

[1] Zeil J, Hofmann M I, Chahl J S. Catchment areas of panoramic snapshots in outdoor scenes[J]. JOSA A, 2003, 20(3): 450-469.

[2] Collett M, Chittka L, Collett T S. Spatial memory in insect navigation. Current Biology, 2013, 23(17): R789-R800.

[3] Wehner R, Michel B, Antonsen P. Visual navigation in insects: coupling of egocentric and geocentric information. Journal of Experimental Biology, 1996, 199(1): 129-140.

[4] Mair E, Augustine M, Jäger B, et al. A biologically inspired navigation concept based on the landmark-tree map for efficient long-distance robot navigation. Advanced Robotics, 2014, 28(5): 289-302.

[5] Arena P, Cruse H, Fortuna L, et al. Adaptive bioinspired landmark identification for navigation control. Bioengineered and Bioinspired Systems III. International Society for Optics and Photonics, 2007, 6592: 65920L.

[6] Stürzl W, Zeil J, Boeddeker N, et al. How wasps acquire and use views for homing[J]. Current Biology, 2016, 26(4): 470-482.

[7] Murray T, Zeil J. Quantifying navigational information: The catchment volumes of panoramic snapshots in outdoor scenes. PloS one, 2017, 12(10): e0187226.

[8] Schofield A J, Gilchrist I D, Bloj M, et al. Understanding images in biological and computer vision. 2018.

[9] Shubham Jain, An Overview of Regularization Techniques in Deep Learning, Analytics Vidhya. 2018.

[10] Ardin, Paul, et al. "Using an insect mushroom body circuit to encode route memory in complex natural environments." PLoS computational biology 12.2 (2016): e1004683.

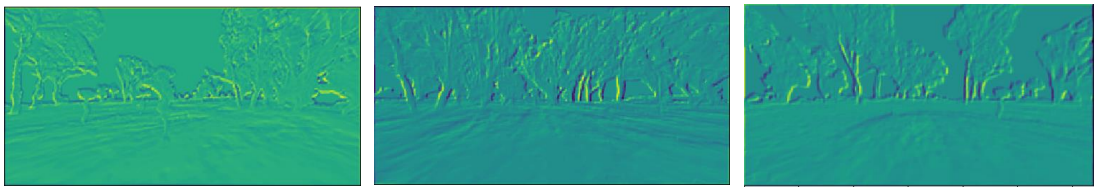
Appendix

Visualization of neuron layer in the CNN

With an input visual image, the trained 11-layer CNN is able to process such spatial information and generate 6-length vector that contribute to the predicted state V_{pred}.

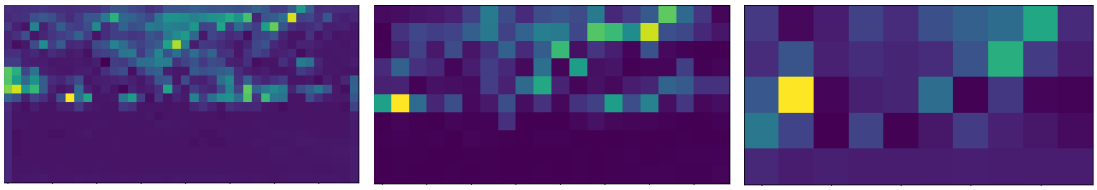
The transmission of information is conducted by these 11 layers with a fixed sequence. The processing of the CNN can be unveil with the mid-product feature maps, and we can notice a functional differentiation of different layers.

a Input visual images of left view, front view and right view



b The 1^{st} layer, spatial feature extraction

c From the 2^{nd} to the 4^{th} layer, visual activation neurons



d From the 5^{nd} to the 7^{th} layer, feature processing neurons

e From the 8^{nd} to the 10^{th} layer, decision-making neurons

Figure 13: The visualization of all feature maps in our CNN

In Fig. 13, the functions of different layers evolve to differentiate after training, and we can find the functions of layers efficiently process the visual information sequentially.

The predicted 6-length vector can be considered as the output neuron of the network, and we illustrate each of these neurons as below. The time-domain response of these neurons also differentiate, while we can observe the motion of an agent with these bottom-level neurons.

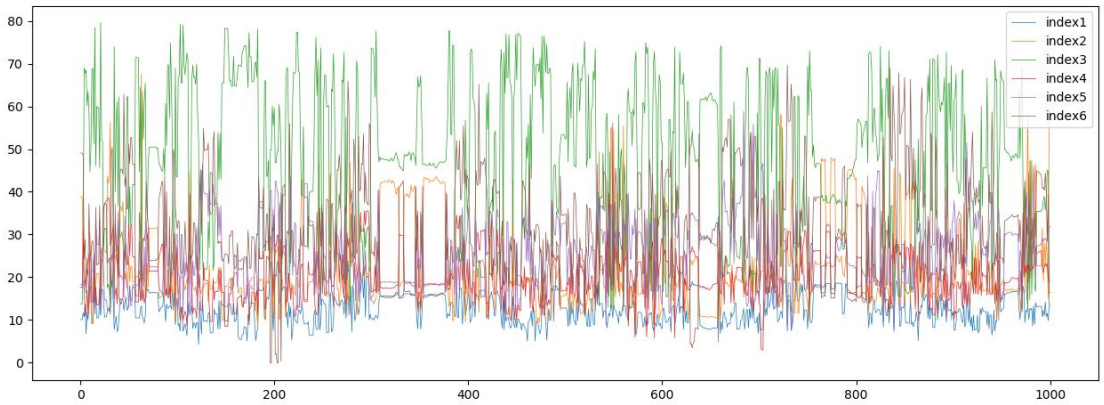


Figure 14: The six decision neuron of the bottom layer in 1000 continuous frames

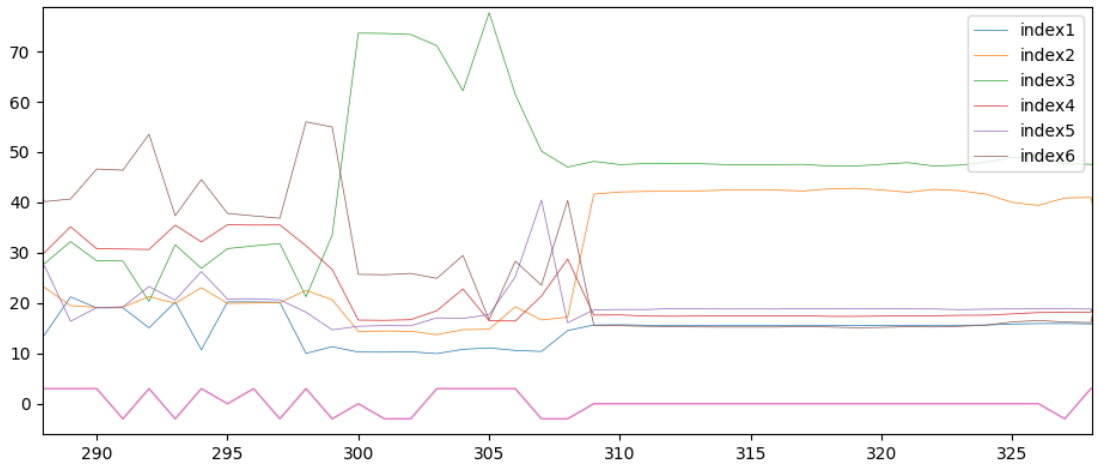


Figure 15: The six decision neuron of the bottom layer from the 290^{th} to the 325^{th} frames. The agent is static between the 310^{th} to the 325^{th} frames, and the neurons show a unchanged neuron response